

Gene Mapping and Phylogenetic Analysis of the Complete Genome from 30 Single-Stranded RNA Male-Specific Coliphages (Family *Leviviridae*)[▽]

Stephanie D. Friedman,^{1,2*} Fred J. Genthner,¹ Jennifer Gentry,² Mark D. Sobsey,² and Jan Vinjé³

U.S. Environmental Protection Agency, Gulf Ecology Division, Gulf Breeze, Florida 32561¹; University of North Carolina at Chapel Hill, Department of Environmental Sciences and Engineering, University of North Carolina Gillings School of Global Public Health, Chapel Hill, North Carolina 27599²; and National Calicivirus Laboratory, Gastroenteritis and Respiratory Viruses Laboratory Branch, Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333³

Received 25 June 2009/Accepted 14 August 2009

Male-specific single-stranded RNA (FRNA) coliphages belong to the family *Leviviridae*. They are classified into two genera (*Levivirus* and *Allolevivirus*), which can be subdivided into four genogroups (genogroups I and II in *Levivirus* and genogroups III and IV in *Allolevivirus*). Relatively few strains have been completely characterized, and hence, a detailed knowledge of this virus family is lacking. In this study, we sequenced and characterized the complete genomes of 19 FRNA strains (10 *Levivirus* strains and 9 *Allolevivirus* strains) and compared them to the 11 complete genome sequences available in GenBank. Nucleotide similarities among strains of *Levivirus* genogroups I and II were 75% to 99% and 83 to 94%, respectively, whereas similarities among strains of *Allolevivirus* genogroups III and IV ranged from 70 to 96% and 75 to 95%, respectively. Although genogroup I strain fr and genogroup III strains MX1 and M11 share only 70 to 78% sequence identity with strains in their respective genogroups, phylogenetic analyses of the complete genome and the individual genes suggest that strain fr should be grouped in *Levivirus* genogroup I and that the MX1 and M11 strains belong in *Allolevivirus* genogroup III. Strains within each genus share >50% sequence identity, whereas between the two genera, strains have <40% nucleotide sequence identity. Overall, amino acid composition, nucleotide similarities, and replicase catalytic domain location contributed to phylogenetic assignments. A conserved eight-nucleotide signature at the 3' end of the genome distinguishes leviviruses (5' ACCACCCA 3') from alloleviviruses (5' TCCTCCCA 3').

Male-specific RNA (FRNA) coliphages are single-stranded RNA (ssRNA) viruses that are found throughout the world in bacterial isolates associated with sewage and feces in mammals (12). They possess a positive-sense genome ranging from 3.8 to 4.2 kb in size enclosed by a nonenveloped 26-nm icosahedral capsid (5). The natural host is restricted to gram-negative bacteria (17) expressing a male factor F⁺, Hfr, or F' (44). For successful infection, the host must possess a fertility (F) sex pilus, coded on the F plasmid of *Escherichia coli* (28), or chromosomal marker Hfr (6), as infection occurs by attachment to this receptor site (7).

FRNA phages belong to the family *Leviviridae*, which is further subdivided into two genera (*Levivirus* and *Allolevivirus*). *Levivirus* is subdivided into genogroups I and II, and *Allolevivirus* is subdivided into genogroups III and IV. Historically, separation into subgroups was based on serological properties (33), sedimentation, density, and molecular weight (36). Recently, genomic data have provided an additional subgrouping tool (41).

Based on a limited number of completely sequenced FRNA genomes, four genes were identified (4). These genes code for an assembly or maturation protein, capsid protein, lysis pro-

tein, and replicase protein in the leviviruses, whereas the lysis protein is replaced by a read-through protein in alloleviviruses. Nonstructural and structural proteins are encoded by the *Leviviridae* viral genome (5). Each FRNA virion contains 1 copy of positive-sense ssRNA, 180 copies of the capsid or coat protein, 1 copy of the assembly or maturation protein, and, in the alloleviviruses, approximately 15 copies of the read-through protein (38, 39, 42).

In this study, the complete genomes of 19 FRNA strains representing the four known genogroups were sequenced and compared to 11 FRNA sequences available in the National Center for Biotechnology Information (NCBI) GenBank, for a total analysis of 30 FRNA genomes. Phylogenetic profiles, nucleotide sequence similarity, amino acid compositions, open reading frame (ORF) positions, and subsequent gene locations were compared. The results of this study will contribute to a better understanding of the ecology of FRNA coliphages as well as provide a more substantial genetic database to design molecular FRNA detection and identification methods.

MATERIALS AND METHODS

FRNA coliphage strains and RNA extraction. FRNA prototype strains used in this study, MS2, GA, Qβ, FI, and SP, were kindly provided by K. Furuse (Tokai University, Japan). Prototype strain fr was provided by A. Boehm (Stanford University, Stanford, CA), and prototype strains MX1 and M11 were obtained from the University of North Carolina, Chapel Hill, collection. FRNA strains ST4, TW18, VK, and BZ1 were a gift from J. van Duin (Leiden University, The Netherlands). Field-collected strains BR1, BR8, and BR12 were generously provided by Brian Robinson (NOAA, Charleston, SC), and strain R17 was

* Corresponding author. Mailing address: U.S. Environmental Protection Agency, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32561. Phone: (850) 934-2468. Fax: (850) 934-9201. E-mail: friedman.stephanie@epa.gov.

[▽] Published ahead of print on 26 August 2009.

obtained from the Felix D'Herelle Reference Centre for Bacterial Viruses, Université Laval, Quebec, Canada. Additional field strains DL1, DL2, DL13, DL16, J20, T72, DL10, DL20, HL4-9, HB-P22, and HB-P24, collected from wastewater, surface waters, swine lagoons, and chicken litter, were used in this study (11). Preliminary subgrouping of all 19 strains was conducted by reverse line blot hybridization (41).

Each strain was plaque purified and further enriched using *Escherichia coli* HS(pFamp)R as the host (41). Aliquots of approximately 1 to 2 ml of the purified viral supernatant were frozen at -75°C .

Coliphage RNA was extracted from purified virus as described by Stewart et al. (32) by using a QIAamp viral RNA minikit (Qiagen, Valencia, CA). Purified RNA was stored frozen at -20°C .

Generating cDNA from polyadenylated RNA. For cDNA synthesis, strain MS2 was used as the positive control. First, viral RNA was 3' polyadenylated with yeast poly(A) polymerase (USB, Inc., Cleveland, OH) and 25 mM ATP in a 50- μl reaction volume (USB). The 50- μl reaction volume was prepared with 10 μl of 5 \times poly(A) polymerase reaction buffer, 10 μl RNA, 2 μl of 25 mM ATP, 0.7 μl of 600 U poly(A) polymerase, and 27.3 μl nuclease-free water. The mixture was incubated at 37°C for 5 min and then placed on ice for enzymatic termination. Polyadenylated RNA was either immediately frozen or used as a template for cDNA synthesis.

Full-length cDNA was prepared using an oligo(dT) reverse primer supplied with the reverse transcriptase MonsterScript 1st Strand cDNA synthesis kit (Epicentre, Madison, WI) as outlined by the manufacturer. The single-stranded cDNA was used as a template for the PCR. To verify the successful generation of full-length cDNA, amplification of a small region of the 5' end of strain MS2 was used as a positive control (19).

Primer walking. To amplify the 1-kb region between the replicase gene and the 3' end of the genome, strain-specific forward primers were designed based on a 200-nucleotide (nt) region of the replicase gene (41) and utilized along with an oligo(dT) reverse primer. To amplify the upstream region of the genome, reverse primers were designed based on the replicase gene sequence of each strain and forward primers were designed based on available FRNA coliphage sequences in GenBank. As sequences were generated (Sequetech, Mountain View, CA), reverse primers were designed to amplify overlapping sections of the genome. The majority of the genome was sequenced by primer walking.

5' amplification of cDNA ends. The nucleotide sequence of the 5' region was determined by rapid amplification of cDNA ends by using a Smart Race cDNA amplification kit (Clontech, Mountain View, CA) with only minor modifications. First-strand cDNA synthesis was carried out on ice with a 250- μl thin-walled PCR tube by combining 3 μl RNA, 1 μl of 10 μM gene-specific reverse primer, and 1 μl Smart oligonucleotide. The 5- μl reaction volume was briefly centrifuged, and the following components were added: 2 μl of 5 \times First Strand buffer (Invitrogen, Carlsbad, CA), 1 μl of 20 mM dithiothreitol, 1 μl of 10 mM deoxynucleoside triphosphate, and 1 μl SuperScript II (Invitrogen, Carlsbad, CA). Following a brief centrifugation, the mixture was incubated for 90 min at 42°C . To dilute the first-strand cDNA, 20 μl of Tricine-EDTA buffer was added, and the mixture was heated for 7 min at 72°C . The reaction generated double-stranded cDNA. The cDNA was frozen at -20°C and used for subsequent PCRs. In all experiments, a spectrophotometer (NanoDrop Technologies, Wilmington, DE) was used to determine nucleic acid concentrations.

Long template PCR, cloning, and sequencing. The cDNA was amplified by using Phusion DNA polymerase (New England Biolabs, Ipswich, MA) in a master mix containing 10 μl of 5 \times Phusion buffer, 0.2 mM deoxynucleoside triphosphate, 1 μl of 10 μM forward primer, 1 μl of 10 μM reverse primer, 3% dimethyl sulfoxide, 2 μl cDNA, and 0.5 μl Phusion *Taq* in a 50- μl reaction volume by using the following cycle parameters: one cycle denaturation at 98°C (1 min) followed by 35 cycles at 98°C (30 s), 48°C (1 min), and 72°C (3 min) followed by a 10-min extension at 72°C . For each reaction, positive controls were prepared using primers MJV82 and JV81 for leviviruses and MJV82 and JV41 for alloviruses (41). A no-template negative control was included.

PCR products were separated by electrophoresis in a 1.5% agarose gel, stained with SYBR Gold nucleic acid gel stain (Molecular Probes, Carlsbad, CA), and visualized under blue light (Dark Reader transilluminators; Clare Chemical Research, Dolores, CO).

Blunt-end PCR products were excised using a gel extraction tool (USA Scientific Plastics, Ocala, FL) and purified according to the instructions of the manufacturer (QuickClean 5 M gel extraction kit; GenScript Corporation, Piscataway, NJ).

Gel-purified DNA was cloned using a Zero Blunt TOPO PCR cloning kit (Invitrogen, Carlsbad, CA). Colonies of transformed *E. coli* cells were screened for positive inserts by using whole-cell PCR and Phusion DNA polymerase as described above with the following cycle modifications: one cycle of denaturation

at 98°C (3 min) followed by 35 cycles at 98°C (10 s), 57°C (30 s), and 72°C (30 s) followed by a 10-min extension at 72°C . Amplicons were separated by electrophoresis in 1.5% agarose gel in 0.5 \times Tris-acetate-EDTA, stained with 20 $\mu\text{g}/\text{ml}$ ethidium bromide, and visualized under UV light (UVP, Upland, CA). Clones with the appropriate-size PCR amplicon were selected for plasmid purification (QIAprep Spin Miniprep kit; Qiagen, Valencia, CA).

Each PCR amplicon was cloned, and three to five clones were sequenced. PCR products from reactions from 5' rapid amplification of cDNA ends were sequenced directly. To achieve publication-quality sequence data, both forward and reverse strands were sequenced (Sequetech, Mountain View, CA).

To avoid contamination, a PCR hood (AirClean 600; AirClean Systems, Raleigh, NC) located in a designated clean room was used to prepare master mixes. PCR amplification, electrophoresis, template, and/or viral preparations (35) were conducted in individual assigned rooms based on designated use.

Sequence analyses. Raw sequences from three to five individual clones were imported and aligned using BioEdit v7.0.1 (14) followed by Basic Local Alignment Search Tool (BLAST; National Center for Biotechnology Information) analyses for sequence and phylogenetic confirmations. Full-length sequences from all strains were aligned with those of prototype strains (GenBank) by using ClustalW.

Similarity analyses were evaluated using SimPlot version 3.5.1 (18). The percent similarity was calculated within a sliding window of 200 bp with a step size of 20 bp between plots.

Amino acid analysis. Deduced amino acid sequences corresponding to each of the four genes were determined using a computer-generated DNA-to-protein translation tool, ExpASY (<http://ca.expasy.org/>). Predicted protein sequence motifs were identified by PROSITE (<http://ca.expasy.org/prosite/>), and protein families and domains were modeled in Pfam (<http://pfam.janelia.org/>). Genetic distance was calculated for each protein within the respective genogroup by ClustalW alignment and neighbor-joining analysis.

Phylogenetic analysis. Sequence data were analyzed using BioNumerics software, version 3.5 (Applied Maths, Saint-Martens-Latem, Belgium). Phylogenetic trees were built by global cluster analysis performed on multiple aligned sequences and clustered by unweighted-pair group method using arithmetic averages. A bootstrap analysis, based on 10,000 substitutions, was used to measure cluster significance. The reliability of each cluster was expressed on a percentage basis.

Nucleotide sequence accession numbers. The accession numbers of some full-length *Leviviridae* sequences (for genogroup I, NC_001417 [MS2], AF195778 [M12], and X15031 [fr]; for genogroup II, NC_001426 [GA] and AF227250 [KU1]; for genogroup III, AF052431 [M11], AY099114 [Q8], and AF059242 [MX1]; and for genogroup IV, X07489 [SP], AF059243 [NL95], and EF068134 [FI]) and a partial sequence of genogroup II strain TL2 (AB218927) were available in GenBank.

The GenBank accession numbers generated in this study are as follows: for genogroup I, EF107159 (DL1), EF108464 (DL16), EF108465 (R17), EF204939 (J20), and EF204940 (ST4); for genogroup II, FJ483837 (DL10), FJ483838 (T72), and FJ483839 (DL20); for genogroup III, FJ483840 (TW18), FJ483841 (HL4-9), FJ483842 (BR12), FJ483843 (VK), and FJ483844 (BZ1); and for genogroup IV, FJ539132 (HB-P22), FJ539133 (HB-P24), FJ539134 (BR1), and FJ539135 (BR8).

RESULTS

Comparison of full-length genome sequences. Full-length genome sequences of 19 FRNA strains were determined in this study and compared to 11 strains previously published in GenBank (Table 1).

Seven genogroup I strains (DL1, DL2, DL13, DL16, ST4, R17, and J20) were sequenced and compared to genogroup I prototype strains MS2, M12, and fr (Table 2). Genogroup I strains DL2 and DL13 were omitted from Table 2, as they were >99% identical, differing by only 4 nt from the DL16 genome. MS2 and ST4 were 98.7% similar to each other. Sequence similarity among genogroup I strains ranged from 75.3 to 98.7%, with strain fr forming a separate subgroup (Table 2; Fig. 1).

Sequences of three genogroup II strains (DL10, DL20, and T72) were compared to the sequences of genogroup II prototype strains GA and KU1. Among genogroup II strains, nucleo-

TABLE 1. *Leviviridae* strain identifications and ORF positions^a

Group and strain	Length in nucleotides (amino acids)					Accession no.
	ORF1	ORF2	ORF3	ORF4	CG	
Genogroup I						
MS2 ^c	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	1,761–3,398 (545)	3,569	NC_001417
fr ^c	129–1,310 (393)	1,336–1,728 (130)	1,691–1,906 (71)	1,762–3,399 (545)	3,575	X15031
M12 ^c	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	ND	3,340 ^b	AF195778
DL1	131–1,312 (393)	1,336–1,728 (130)	1,679–1,906 (75)	1,762–3,399 (545)	3,570	EF107159
DL2 ^d	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	1,761–3,398 (545)	3,491 ^b	NA
DL13 ^d	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	1,761–3,398 (545)	3,491 ^b	NA
DL16	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	1,761–3,398 (545)	3,569	EF108464
J20	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	1,761–3,398 (545)	3,569	EF204939
ST4	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	1,761–3,398 (545)	3,569	EF204940
R17	130–1,311 (393)	1,335–1,727 (130)	1,678–1,905 (75)	1,761–3,398 (545)	3,569	EF108465
Genogroup II						
GA ^c	136–1,308 (390)	1,325–1,717 (130)	1,717–1,908 (63)	1,749–3,327 (532)	3,466	NC_001426
KU1 ^c	137–1,309 (390)	1,325–1,683 (130)	1,735–1,929 (64)	1,770–3,348 (532)	3,486	AF227250
T72 ^d	137–1,309 (390)	1,325–1,683 (130)	1,735–1,929 (64)	1,770–3,348 (532)	3,393 ^b	FJ483838
DL10 ^d	136–1,308 (390)	1,325–1,717 (130)	1,717–1,908 (63)	1,749–3,327 (532)	3,412 ^b	FJ483837
DL20	129–1,301 (390)	1,318–1,710 (130)	1,710–1,901 (63)	1,742–3,320 (532)	3,458	FJ483839
Genogroup III						
Qβ ^{c,d}	62–1,324 (420)	1,345–1,746 (133)	1,345–2,334 (328)	2,353–4,122 (592)	4,160 ^b	AY099114
M11 ^c	57–1,322 (421)	1,344–1,744 (133)	1,344–2,333 (328)	2,352–4,118 (588)	4,217	AF052431
MX1 ^c	56–1,321 (421)	1,343–1,744 (133)	1,343–2,332 (328)	2,351–4,111 (586)	4,215	AF059242
TW18	62–1,324 (420)	1,345–1,746 (133)	1,345–2,334 (328)	2,353–4,122 (592)	4,218	FJ483840
HL4-9	62–1,324 (420)	1,345–1,746 (133)	1,345–2,337 (329)	2,356–4,125 (592)	4,221	FJ483841
BR12	62–1,324 (420)	1,345–1,746 (133)	1,345–2,334 (328)	2,353–4,122 (592)	4,218	FJ483842
VK	62–1,324 (420)	1,345–1,746 (133)	1,345–2,334 (328)	2,353–4,122 (592)	4,218	FJ483843
BZ1	62–1,324 (420)	1,346–1,747 (133)	1,346–2,335 (328)	2,354–4,123 (592)	4,219	FJ483844
Genogroup IV						
NL95 ^c	53–1,371 (442)	1,402–1,800 (132)	1,402–2,394 (331)	2,414–4,144 (576)	4,248	AF059243
SP ^c	55–1,400 (450)	1,427–1,825 (132)	1,427–2,422 (330)	2,442–4,172 (576)	4,276	X07489
FI ^{c,d}	55–1,371 (438)	1,392–1,791 (132)	1,392–2,391 (332)	2,406–4,167 (586)	4,184 ^b	EF068134
HP-P22	52–1,367 (440)	1,395–1,793 (132)	1,395–2,387 (329)	2,407–4,137 (576)	4,241	FJ539132
HP-P24	53–1,368 (441)	1,397–1,795 (132)	1,397–2,389 (329)	2,409–4,139 (576)	4,243	FJ539133
BR1	52–1,397 (450)	1,424–1,822 (132)	1,424–2,419 (330)	2,439–4,169 (576)	4,273	FJ539134
BR8	52–1,397 (450)	1,424–1,822 (132)	1,424–2,419 (330)	2,439–4,169 (576)	4,273	FJ539135

^a The number of amino acids for each gene is in parentheses. Proteins were determined by mapping of ORFs (BioEdit) and translation of nucleotide sequences to amino acids by using the ExPASy (<http://ca.expasy.org/>) DNA-to-protein translation tool. CG, complete genome; ND, not determined; NA, not applicable.

^b Nearly full-length genome.

^c Previously published GenBank genome.

^d ORF nucleotide position based on alignment.

tide sequence similarities ranged from 83.3 to 93.8%, with strains DL10, DL20, and GA having the highest sequence identities (93.4 to 93.7%), whereas strains T72 and KU1 formed a separate subcluster (Table 2; Fig. 1). Strains in genogroup I had only 50% sequence similarity (range of 46.7 to 53.9%) with strains in genogroup II (Table 2; Fig. 2A). Interestingly, all *Levivirus* strains (genogroups I and II) had an identical 8-nt sequence at the 3' terminus, 5' ACCACCCA 3' (Table 3).

Allolevivirus genogroup III strains formed two different subclusters (Fig. 1). Strains VK, HL4-9, BR12, BZ1, TW18, and Qβ, having nucleotide sequence similarities ranging from 91.9 to 95.7% (Table 2), formed the first subcluster (Fig. 1). The second subcluster was formed with genogroup III strains MX1 and M11, as these strains shared 87% nucleotide similarity. The nucleotide similarities of strains between the two genogroup III subclusters ranged from 69.8 to 71.3% (Table 2). Genogroup III strains shared <40% sequence identity (29.7 to 39.1%) with *Levivirus* genogroups I and II (Table 2). All

Allolevivirus strains had an identical 3' terminus signature sequence of 5' TCCTCCCA 3' (Table 3).

Sequences of genogroup IV strains BR1, BR8, HB-P22, and HB-P24 were compared to the sequences of genogroup IV prototype strains SP, FI, and NL95. Genogroup IV *Allolevivirus* strains shared nucleotide sequence identities ranging from 74.9 to 95.0%, with the closest identities being 95.0% between strains BR8 and BR1 (Table 2). Strain HB-P24 shared 90.2% nucleotide identity with prototype strain NL95, whereas strains BR8 and BR1 grouped with prototype strain SP (91.1 to 91.7%). In contrast, strain FI formed a unique phylogenetic subcluster (Fig. 1). Genogroup IV nucleotide sequence identity was 53.5 to 57.9% with *Allolevivirus* genogroup III (Table 2; Fig. 2B) and <40% (31.8 to 38.7%) with *Levivirus* genogroups I and II.

ORF analyses. (i) *Levivirus*. In *Levivirus* genogroup I, the ORF start and stop codons were located at nucleotide positions identical or very similar to those previously reported for strain MS2 (9). With the exceptions of strain DL1 and fr, all

TABLE 2. *Leviviridae* nucleotide percent similarity

Strain no./name/ genogroup	% Nucleotide sequence similarity with the indicated strain																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1/MS2/I	100.00																											
2/ST4/I	98.71	100.00																										
3/R17/I	96.39	96.67	100.00																									
4/M12/I	92.78	92.93	92.93	100.00																								
5/J20/I	92.19	92.47	92.38	91.82	100.00																							
6/DL1/I	92.16	92.41	92.16	92.48	93.95	100.00																						
7/DL16/I ^a	91.68	92.07	91.96	92.00	94.73	94.79	100.00																					
8/fr/I	75.27	77.09	75.89	75.43	76.63	76.55	75.65	100.00																				
9/DL10/II	53.13	52.18	52.54	52.29	53.47	53.15	52.35	49.93	100.00																			
10/DL20/II	52.89	52.23	52.17	52.18	53.85	50.65	53.69	51.68	93.84	100.00																		
11/T72/II	52.37	52.35	52.49	50.41	51.42	49.99	49.93	46.74	85.97	85.17	100.00																	
12/GA/II	51.35	51.44	51.76	51.75	51.14	50.17	51.79	48.25	93.67	93.43	84.79	100.00																
13/KU/II	49.55	49.61	49.69	49.72	51.15	50.23	49.58	48.71	84.69	83.83	89.89	83.30	100.00															
14/VK/III	38.66	34.85	37.90	33.79	34.59	37.31	36.13	33.01	37.39	37.37	36.40	37.61	36.39	100.00														
15/HL4-9/III	37.55	36.25	37.05	36.19	35.25	32.75	36.68	36.13	34.16	38.11	36.51	34.38	39.25	91.97	100.00													
16/Qβ/III	36.87	36.33	36.94	36.15	38.27	34.07	37.88	32.05	38.02	38.44	36.58	34.85	36.62	92.09	94.76	100.00												
17/BR1/III	36.85	36.91	36.91	36.47	37.17	36.32	36.19	32.69	37.95	38.49	36.83	38.73	36.44	95.69	91.87	91.97	100.00											
18/BZ1/III	36.79	36.91	37.61	36.31	36.68	37.05	36.53	37.00	31.95	32.27	36.70	33.04	39.06	93.37	92.30	92.24	92.87	100.00										
19/TW18/III	36.39	36.73	36.39	35.97	35.28	32.71	36.19	31.71	37.63	38.22	36.13	34.73	36.31	92.25	95.90	95.58	92.06	92.65	100.00									
20/SP/IV	36.29	34.64	34.73	34.64	34.90	34.77	34.18	33.77	37.09	36.69	32.80	37.13	32.32	55.49	57.75	54.19	54.64	53.09	55.92	100.00								
21/HB-P24/IV	35.39	35.47	34.70	35.43	37.00	33.64	38.54	35.85	37.30	33.03	37.54	36.81	38.73	56.13	55.93	53.53	56.73	54.87	55.89	79.37	100.00							
22/BR1/IV	34.13	34.13	36.77	34.61	34.59	37.12	34.30	37.55	37.21	33.08	32.77	33.18	31.81	53.61	57.37	53.93	55.36	54.31	56.94	91.73	79.80	100.00						
23/FI/IV	34.05	33.96	32.90	37.03	33.81	37.93	37.63	33.20	31.99	33.41	33.07	33.30	32.53	54.43	56.67	55.65	54.85	55.98	56.15	76.63	78.95	74.90	100.00					
24/BR8/IV	34.04	36.61	34.19	33.43	34.70	36.68	34.76	33.66	36.66	36.19	36.83	36.91	36.92	55.27	56.64	53.17	55.33	55.49	54.82	91.05	78.87	95.03	74.57	100.00				
25/M11/IV	33.47	31.11	33.49	30.54	35.66	32.79	32.95	36.14	33.59	33.47	31.82	33.75	30.57	71.33	70.43	70.45	71.05	70.66	71.23	53.67	55.84	52.33	56.03	51.61	100.00			
26/NL95/IV	33.10	37.11	32.64	32.55	36.94	32.75	32.87	33.46	36.05	36.04	37.76	37.58	36.99	55.99	56.39	55.81	55.80	54.64	53.92	78.50	90.22	78.47	76.07	78.32	55.89	100.00		
27/MX1/IV	32.53	32.33	31.21	29.73	36.44	37.49	32.27	33.09	33.47	32.59	37.81	33.45	32.81	70.27	70.55	69.90	70.13	69.77	70.92	52.00	55.08	52.46	54.03	51.19	87.15	54.77	100.00	
28/HB-P22/IV	32.30	32.87	32.39	32.28	33.43	32.55	32.65	33.06	36.29	37.01	37.56	36.69	37.54	57.79	57.99	57.37	57.65	56.32	57.34	81.80	82.83	82.07	77.69	81.38	54.97	82.25	54.43	100.00

^a DL2 and DL13 were omitted. DL16 was used as a representative strain, as all three strains differed by only 4 nt.

MS2-like strains (DL1, DL2, DL13, DL16, J20, ST4, R17, and M12) had their four ORFs located at similar nucleotide positions (Table 1). AUG was found to be the start codon for all four ORFs within genogroup I with two exceptions: (i) MS2, ST4, and fr had GUG as the start codon for ORF1, and (ii) the strain fr lysis gene start codon was UUG (Table 4).

The nucleotide start and stop positions of genogroup II genes of strains T72 and KU1 were similar to each other, whereas the nucleotide stop and start positions of strains DL10 and DL20 were similar to those of strain GA (Tables 1 and 4). The start codon for all ORFs in genogroup II strains GA, DL10, and DL20 was AUG (Table 4). The lysis gene start codon in strains T72 and KU1 was UUG (Table 4). In addition, there was an 18-nt insertion between the capsid gene stop codon (ORF2) and the lysis gene start codon (ORF3) in strains T72 and KU1 that was absent in the other genogroup II strains. However, a translation coupled stop-start codon (UAAUG) was observed in the lysis protein for genogroup II strains GA, DL10, and DL20.

(ii) *Allolevivirus*. The *Allolevivirus* genome possesses four genes and three start codons, as the capsid and read-through genes share a single ORF (ORF2/ORF3) (Table 4). The ORF alignment positions for all genogroup III strains except MX1 and M11 were very similar if not identical. Although ORFs of strain Qβ aligned perfectly with those of the other genogroup III strains, the GenBank-acquired Qβ sequences were not complete. Thus, individually mapped ORF positions varied slightly (Table 1). ORF and Shine-Dalgarno positions of the assembly, capsid, and read-through genes of MX1 and M11 were similar to those of the other genogroup III strains, but the replicase gene ORF and Shine-Dalgarno positions differed (Table 4).

Within genogroup IV, phages BR1, BR8, NL95, and SP possessed similar ORF positions, whereas strains HB-P22, HB-P24, and FI possessed similar ORF nucleotide positions (Tables 1 and 4).

Shine-Dalgarno sequences. In prokaryotes, the Shine-Dalgarno core consensus sequence (GGAGG) or slight variations of the core sequence are located upstream from the ORF start codon (20, 30). Since variations of the core Shine-Dalgarno sequences were observed in this study, spacing between the start codon and Shine-Dalgarno sequence was defined as the number of nucleotides between the last base of the Shine-Dalgarno sequence and the first base of the start codon. For *Levivirus* groups I and II, Shine-Dalgarno sequences were located within 4 to 8 nt upstream from ORF1 and ORF2, 8 to 15 nt upstream from ORF3, and 6 nt upstream from ORF4. Shine-Dalgarno sequences for *Allolevivirus* groups III and IV were located 4 or 5 nt upstream from ORF1, 11 nt upstream from ORF2/ORF3, and 5 to 8 nt upstream from ORF4 (Table 4).

Amino acid composition of *Levivirus*. The maturation/assembly proteins in genogroups I and II were 393 and 390 amino acids in length, respectively, and the *Levivirus* capsid protein was 130 amino acids in length (Table 1). The lysis protein of strain fr consisted of 71 amino acids, whereas the remaining genogroup I strains had a lysis protein 75 amino acids in length. An amino acid deletion was observed in the genogroup II lysis protein of strains DL10, DL20, TL2, and GA relative to T72 and KU1 (Table 1). The conserved ssRNA YGDD sequence was located in the replicase protein, and the lengths of these proteins in genogroups I and II were 545 and 532 amino acids, respectively.

Pfam for *Levivirus*. Pfam examines species distribution of the specified protein and to which family or families the protein structure belongs (10). *Levivirus* capsid amino acid compositions were placed into the "Levi_coat" domain, which included all *Leviviridae* strains from genogroups I, II, III, and IV and bacteriophage PRR1 in the Pfam species tree. Pfam searches of genogroup I and II replicase proteins resulted in phages from the *Leviviridae* family and the addition of bacteriophages PRR1, ZR, and BO1, as well as *Acinetobacter* phage

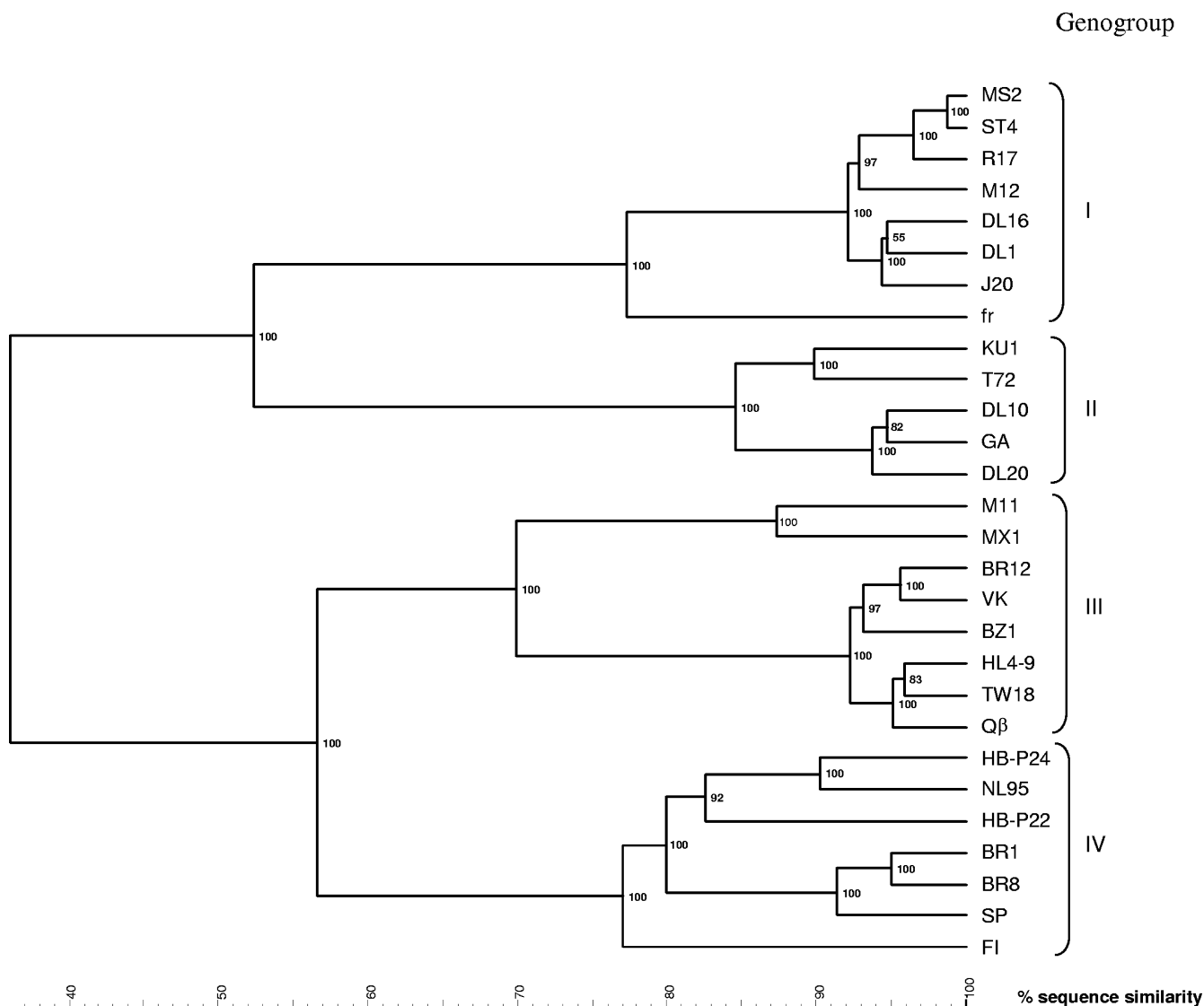


FIG. 1. Unrooted phylogenetic analysis of full-length nucleotide sequences. Nucleotide percent similarity between each male-specific FRNA strain (family *Leviviridae*) is represented on the horizontal axis. The numbers on each node indicate bootstrap values expressed as percentages.

AP205. The replicase protein was placed into the “RNA replicase, beta-chain” domain. The genogroup I lysis protein was not sorted into a family or domain in a PfamA search. A subsequent PfamB search for the genogroup I lysis protein linked it to a lysis domain, and the results matched *Levivirus* genogroup I strains fr, M12, MS2, and JP501. A PfamA maturation protein search generated the “phage_mat-A” domain along with a Pfam species tree including all *Leviviridae* strains plus three additional bacteriophages, PRR1, PP7, and AP205.

Protein sequence motifs of *Levivirus*. Predicted protein motifs, casein kinase II phosphorylation, cyclic AMP (cAMP)- and cGMP-dependent protein kinase phosphorylation, protein kinase C phosphorylation, N myristoylation, N glycosylation, and tyrosine kinase phosphorylation, occurred frequently in the FRNA coliphages. Unique to strain fr was the presence of a leucine zipper in the lysis protein and an amidation motif in the replicase region. The replicase gene RNA-dependent RNA

polymerase catalytic domain occurred at amino acid positions 243 to 373 and 245 to 375 for groups I and II, respectively. Common to every genogroup II strain was a prenyl group binding site (CAAX box) at amino acid positions 529 to 532 in the replicase region.

Genetic distances of *Levivirus*. Excluding that of strain fr, genogroup I amino acid compositions were very conserved, as the genetic distances were small (data not shown). The capsid protein was the most conserved (distance of 0.0000 to 0.0411), followed by maturation protein (0.0046 to 0.0889), replicase (0.0033 to 0.0887), and lysis protein (0.0000 to 0.3416). The capsid protein was identical among strains DL1, DL2, DL13, DL16, and J20. As demonstrated by the distance values (0.2316 to 0.5685), the amino acid compositions of all four proteins in strain fr were not conserved compared to those of the other genogroup I strains.

In genogroup II strains, the capsid protein was the most

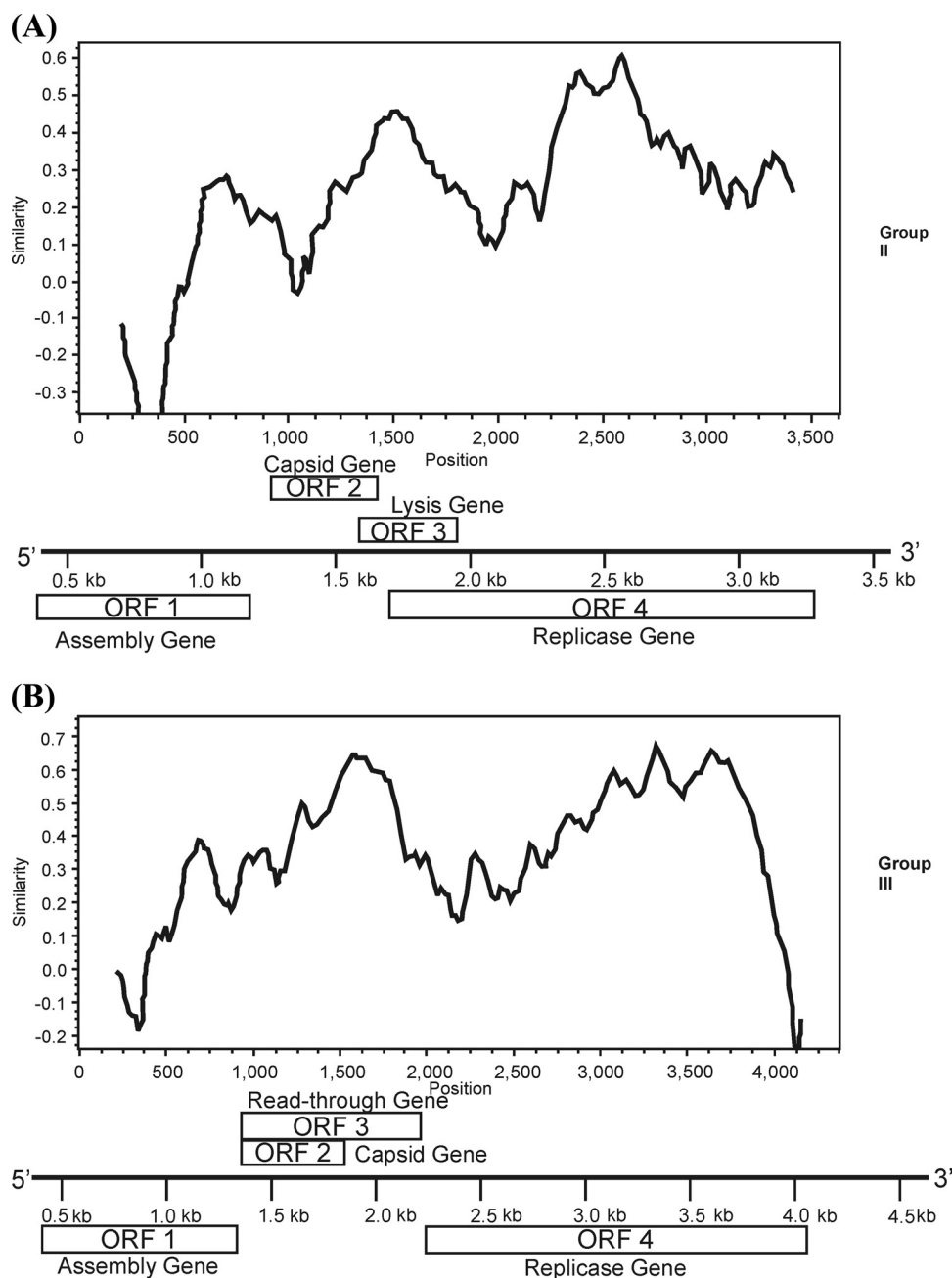


FIG. 2. SimPlot nucleotide similarity and genome organization. The vertical axes show percent similarities within a sliding window 200 bp wide and a step size between plots of 20 bp. The horizontal axes show the approximate nucleotide positions on the genome. (A) *Levivirus* genogroup I queried against genogroup II. (B) *Allovivivirus* genogroup III queried against genogroup IV.

conserved, with a genetic distance of 0.0135 to 0.1116, followed by the lysis (0.0000 to 0.1521), replicase (0.0415 to 0.2160), and maturation (0.0235 to 0.2036) proteins.

Amino acid composition of *Allovivivirus*. The length of the maturation proteins of genogroups III and IV ranged from 420 to 450 amino acids (Table 1). Genogroup IV maturation protein in strains HB-P22, HB-P24, NL95, and FI had a nine-amino-acid deletion compared to the other genogroup IV strains. The lengths of the capsid proteins were 133 and 132 for genogroups III and IV, respectively. Read-through proteins

were 328 to 329 and 329 to 332 amino acids in length for genogroups III and IV, respectively. The replicase was 576 to 592 amino acids in length and contained the YGDD sequence.

Pfam for *Allovivivirus*. Similar to the case for *Levivirus*, a *Allovivivirus* maturation protein search generated the “phage_mat-A” domain, which matched all four genogroups of *Leviviridae* phages described in this study plus the non-FRNA bacteriophage strains PRR1, PP7, and AP205. The PfamA capsid protein search resulted in the family “Levi_coat” and matched all four genogroups in the *Leviviridae* family, along

TABLE 3. Characteristics of *Levivirus* and *Allolevivirus*

Parameter	<i>Levivirus</i>	<i>Allolevivirus</i>
Genome size (nt)	3,458–3,575	4,215–4,276
No. of genes	4	4
No. of proteins	4	4
No. of ORF initiation sites	4	3
Protein names	Assembly/maturation, capsid, lysis, replicase	Assembly/maturation, capsid, read-through, replicase
ORF initiation sites	Single start codon for each gene	Same for second and third genes
Replicase gene	Start codon occurs within lysis gene	Single gene
3' signature	5' ACCACCCA 3'	5' TCCTCCCA 3'

with FRNA phages ZR, TH1, TL2, SD, f2, and BO1 plus the *Pseudomonas* bacteriophage PRR1. Non-FRNA strains PP7 and AP205 were not detected in the capsid search results.

Read-through proteins were grouped as “A1-protein coat readthrough” with PfamB, generating a five-member FRNA strain match of SP, Qβ, NL95, MX1, and M11. As with the *Levivirus* protein, the *Allolevivirus* replicase protein was sorted into the “RNA replicase, beta-chain” family, which included *Leviviridae* strains of all four genogroups with additional FRNA strains ZR and BO1 plus non-FRNA bacteriophages PRR1, PP7, and AP205.

Protein sequence motifs of *Allolevivirus*. As observed for the *Levivirus* genus, the most prevalent protein motifs for *Allolevivirus* were casein kinase II phosphorylation, cAMP- and cGMP-dependent protein kinase phosphorylation, protein kinase C phosphorylation, N myristoylation, N glycosylation, and

tyrosine kinase phosphorylation motifs. With the exception of genogroup III strains MX1 and M11 and genogroup IV strain HB-P24, a cell attachment motif (RGD) was present in the maturation protein. Genogroup IV strains SP, BR8, BR1, and HB-P22 had an additional cell attachment motif in the read-through protein.

The catalytic domain of the RNA-dependent RNA polymerase (replicase protein) was located at amino acid positions 262 to 394 in genogroup III strains, with the exception of strains M11 and MX1. The M11 and MX1 catalytic domain was located at amino acid positions 259 to 391. The catalytic domain of genogroup IV was located at amino acid positions 259 to 391.

Genetic distances of *Allolevivirus*. Genetic distances (data not shown) corresponding to genogroup III proteins were most conserved in the capsid (distance, 0.0000 to 0.3734) followed

TABLE 4. Start codons and Shine-Dalgarno sequences for each FRNA (family *Leviviridae*) gene^a

Group and ORF	Strain	Gene	Sequence (5'–3')
<i>Levivirus</i> genogroup I			
1	All I except fr	Assembly	CCUAGGAGGUUUGACYRUGCGAGC
	fr	Assembly	GCUAGGGAGCCUCGUGUGCGAAAGU
2	All I except fr	Capsid	AACCGGAGUUYGAAGCAUGGCUUCU
	fr	Capsid	CCGAAGGGAGAGCCACAUGGCUUCG
3	All I except fr	Lysis	UGCAAGGUCUCCURAAAGAUGGAA
	fr	Lysis	AACUGGUAACCCAAUUGCAACAGC
4	All I except fr	Replicase	CAUGAGGAUUACCCAUGUCGA
	fr	Replicase	CAUGAGGAAUACCCAUGUCA
<i>Levivirus</i> genogroup II			
1	All II	Assembly	AUACCGGAGGADCUAUGUUUCCGA
2	All II	Capsid	WWAYGGAGUUAGCCAYAUGGCAAC
3	T72, KU1	Lysis	GGCGUAGUUCUUCAGCAUUGGGUC
	DL10, DL20, GA	Lysis	CDCAGAGYGGCUUCUACGCGUAAUGGGUC
4	All II	Replicase	CAUAAGGAAAACCUAUGUUCGUAUUA
<i>Allolevivirus</i> genogroup III			
1	All III	Assembly	DRGAGMMAYAUGCCWM
2/3	All III	Capsid/readthr	UGGGUCAAUUHGAUCAUGGCWAAA
4	All except MX1 and M11	Replicase	AGUAACUAAGGAUGAAAUGCAUGUCUA
	MX1 and M11	Replicase	AGUAACURAAGGAGAUUCGCAUGUCWA
<i>Allolevivirus</i> genogroup IV			
1	All IV	Assembly	CUACAGAGGAGAAUCUAUGCC
2/3	BR1, BR8, NL95, SP	Capsid/readthr	CUUUGGGUCAAUUYGAUCAUGGCAA
	HB-P22, HB-P24, FI	Capsid/readthr	YUUUGGGUCAAUUYGAUCAUGGCWA
4	BR1, BR8, NL95, SP	Replicase	CUUAARRGAGRWAGCAUGYCAA
	HB-P22, HB-P24, FI	Replicase	UYAAAGGAGWUWGCAUGUCUA

^a Start codons (boldface and underlined) and Shine-Dalgarno sequences (underlined) for each FRNA (family *Leviviridae*) gene are presented. The alignments of strains DL1, DL2, DL13, DL16, J20, ST4, R17, M12, and MS2 from *Levivirus* genogroup I are shown, with data for strain fr shown separately. The alignments of strains T72, DL10, DL20, KU1, and GA from *Levivirus* genogroup II, strains TW18, HL4-9, BR12, BZ1, VK, Qβ, MX1, and M11 from *Allolevivirus* genogroup III, and strains HB-P22, HB-P24, BR1, BR8, SP, NL95, and FI from *Allolevivirus* genogroup IV are also shown. readthr, read-through.

by read-through protein (0.0444 to 0.5128) and replicase (0.0278 to 0.6571), with the greatest genetic distance occurring in the maturation protein (0.0347 to 0.8289). Strains BR12 and VK shared identical capsid proteins (distance of 0.0000).

In genogroup IV, the most similar amino acid compositions were found in the capsid (0.0535 to 0.2569), followed by the replicase (distance 0.0474 to 0.3382) and the read-through protein (0.0555 to 0.5072). The greatest genetic distance was observed in the maturation protein (0.0607 to 0.5646).

Phylogenetic analyses. Phylogenetic trees from nucleotide sequences for genogroup I strains clustered into two branches, one branch with nine strains clustered as MS2-like and a second branch with strain fr (Fig. 1). For genogroup II nucleotide sequences, strains KU1 and T72 formed one branch and strains DL10, DL20, and GA formed a second branch. Genogroup III nucleotide sequences clustered into two branches, one branch containing MX1 and M11 and a second branch with Q β -like strains BR12, VK, BZ1, HL4-9, TW18, and prototype Q β . Nucleotide sequence analysis revealed three branches in genogroup IV strains as follows: (i) HB-P24, HB-P22, and prototype NL95, (ii) BR1, BR8, and prototype SP, and (iii) prototype FI.

Overall, phylogenetic trees from amino acid sequences of individual proteins were similar to nucleotide phylogenetic clustering (Fig. 3). For example, each genogroup III protein formed two branches, (i) MX1 and M11 and (ii) Q β -like (six strains). Genogroup IV generated three or four subclusters for each individual protein tree.

Nucleotide SimPlot analysis for all genogroup I and II strains showed that the replicase genes were most similar (approximately 0.5 to 0.6, or 50 to 60% similarity), whereas the assembly or maturation genes were the most dissimilar (<10%) (Fig. 2A). When full-length nucleotide sequences were compared between genogroups III and IV, SimPlot graphs showed similar regions in the capsid (approximately 60%) and the 5' portion of the replicase (approximately 60%) but increased dissimilarity in the assembly and 3' region of the replicase (Fig. 2B).

DISCUSSION

We report the characterization of 30 complete FRNA genomes, 19 of which were newly sequenced in this study. Phylogenetic analyses confirmed that the family *Leviviridae* contains two genera, each with two distinct genogroups. In some cases, the genetic similarity of FRNA strains collected from different continents was greater than 90% within each genogroup, which suggests that geography does not play a significant role in sequence variability (37). The sequences showed great uniformity (i.e., protein length, ORF positions, and replicase catalytic domain) throughout the *Leviviridae* family (15). For example, the number of amino acids in the capsid or coat protein is highly conserved, ranging from 130 to 133 amino acids, indicating that the size of this protein is apparently constant in *Leviviridae* phages (15) and critical for the structure of proper capsid configuration (25).

Unlike the case for genogroup II, the genogroup I lysis gene initiation codon was embedded at the 3' end of the capsid gene. In both genogroups I and II, the lysis gene was out of frame from the capsid gene and the termination codon was

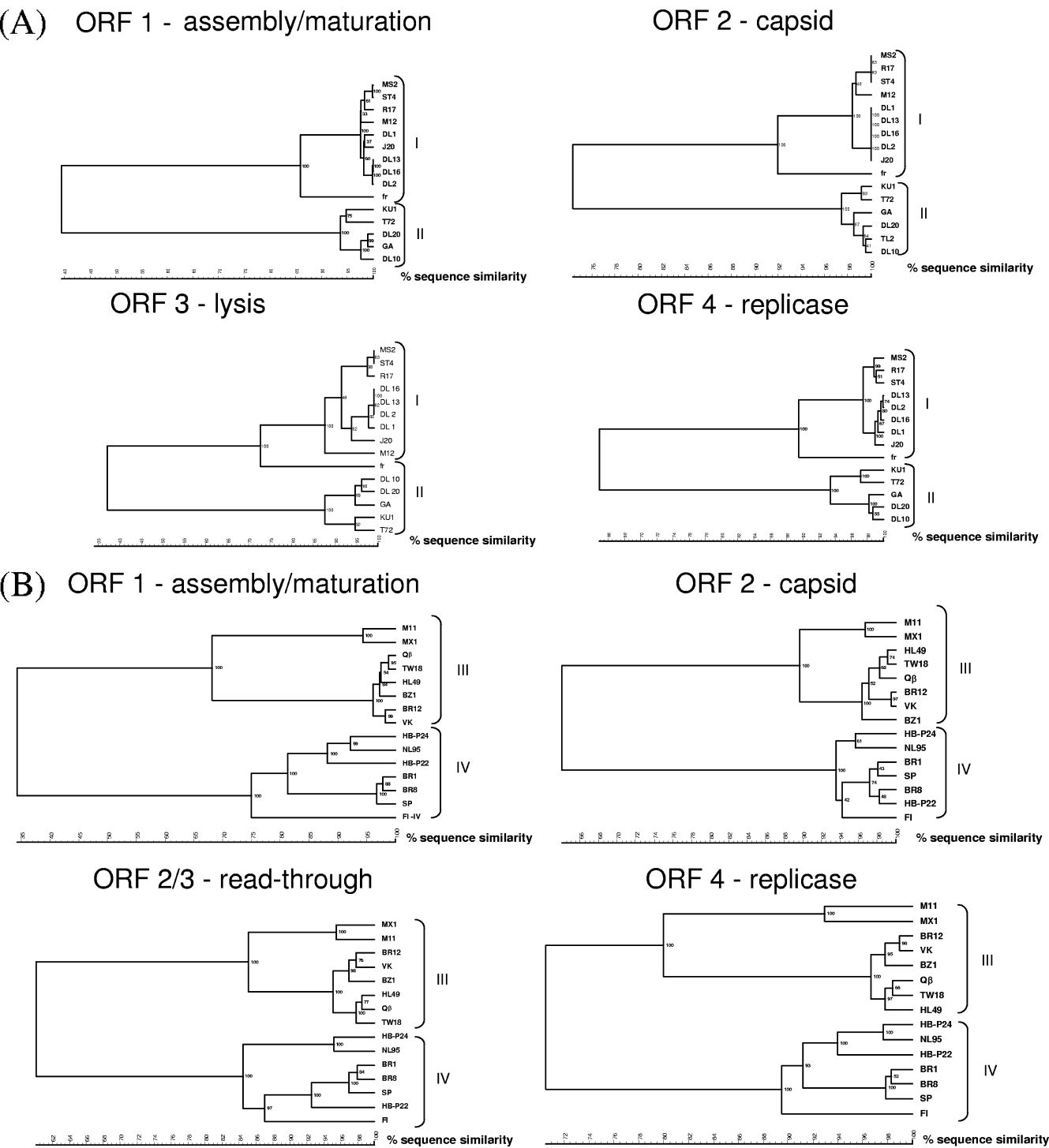
located at the 5' region of the replicase gene. Subsequently, the replicase ORF4 initiation codon was embedded in the lysis gene and also reads out of frame from the lysis gene.

Notably, the genogroup II lysis gene was unique within the *Leviviridae* viruses. Genogroup II strains DL10, DL20, and GA had an overlapping or translational coupled stop-start codon (UAAUG) (16) in the lysis gene, which was not observed in other *Leviviridae* strains. In this case, the frameshifted lysis start codon was coupled to the capsid stop codon. In comparison, the coupled stop-start codon was disrupted between the A and U by an 18-nt insertion in strain T72 and resulted in 17 nt between the capsid stop codon lying out of frame with the lysis start codon. The sequences of strain T72 confirmed an earlier observation in which strain KU1 also had this 18-nt insert (13). In addition, the UUG lysis start codon (13) was unique to strains T72 and KU1 and has been reported to occur at a rate of approximately 3% as an alternative start codon in prokaryotes (3). Despite the dissimilarities in the genogroup II lysis start codon, the translated lysis protein was conserved (93.8 to 100% amino acid similarity), supporting the findings that the variability of the start codon was not involved in gene regulation (13).

All 30 FRNA strains as well as bacteriophages PP7, AP205, and PRR1 contained the replicase YGDD motif, which is conserved among all positive-sense ssRNA viruses (27). Interestingly, all *Leviviridae* genomic sequences possessed a conserved 8-nt signature at the 3' end of the genome, which distinguishes alloviruses (5' TCCTCCCA 3') from leviviruses (5' ACCACCCA 3'). This observation confirms and extends previous reports on a CCCA stretch at the 3' end of the genome (15).

The Shine-Dalgarno sequence, the initiation codon spacing, and the secondary RNA structure function to initiate protein translation by aligning the ribosome with the start codon (30). In this study, Shine-Dalgarno sequences of FRNA viruses were located 4 to 15 nt upstream from the start codon(s). In comparison to prokaryotes, greater than 80% of these Shine-Dalgarno sequences occur within 5 to 13 bases upstream (20).

Stop codons, UAG, UAA, and UGA, serve as signals for peptide chain termination. During translation of the viral RNA coat protein cistron, the UGA stop codon can be read through, resulting in an additional translated product (43). In *Allolevivirus*, a read-through protein is translated when a leaky UGA stop codon is misread as a tryptophan codon (UGG) (39), thereby influencing regulatory control and efficiency of gene expression (1). In comparison, alignment of genogroup III viruses revealed that the *Allolevivirus* maturation protein stop codon is also a UGA; however, in this instance, it is a nonleaky codon. This may occur because the 5' and 3' codons flanking the UGA stop codon influence translation termination efficiency (2, 23, 31). A study on genogroup III strain Q β proposed that programmed read-through was regulated by the 3' nucleotides, specifically an A nucleotide, flanking the stop codon in strain Q β (8). However, alignment of genogroup III and IV nucleotide data in the present study did not reveal a 3' flanking pattern downstream of the UGA stop codon, nor was a 3' nucleotide observed immediately following the stop codon in all genogroup III strains. Noticeably, Q β -like strains contained the 3' A; however, the 3' A was absent in genogroup III strains MX1 and M11. Interestingly, in all *Allolevivirus* se-



quences, a 5' pattern emerged at the read-through UGA stop codon but was absent in the maturation gene UGA stop codon. Beginning 12 nt upstream from the read-through UGA stop codon, the sequence AAY CCR GCR UAY UGA in geno-group III and AAY CCW GCN UAC UGA in genogroup IV

was observed. Nucleotides present in both genogroups III and IV are underlined; these nucleotide triplets coded for the amino acids LNPAY. These findings suggest the upstream sequences may reduce translation termination efficiency of the UGA read-through stop codon in *Allovivivirus* spp.

A cell attachment motif, Arg-Gly-Asp (RGD), was identified in the maturation and/or read-through proteins in the majority of *Allolevivirus* strains but was absent in *Levivivirus* strains. The function of the RGD motif in FRNA coliphages has yet to be experimentally demonstrated but may explain the fact that *Levivivirus* strains attach to the host's pili via the maturation protein, whereas in *Allolevivirus* strains, both the maturation and read-through proteins are required for phage infection (37). For example, the RGD motif was shown to be involved in cell-to-cell adhesion in ssRNA viruses, such as the passaged foot-and-mouth disease virus (21), enterovirus, echovirus 9 strain Barty, coxsackievirus A9, echovirus 22 (24), and bluetongue virus (34). In nearly all astroviruses, an RGD or similar integrin-recognition motif has been identified (40).

In previous reports, ssRNA bacteriophages PP7, PRR1, and AP205 have been compared to FRNA phages. *Pseudomonas aeruginosa* ssRNA phage PP7 shares secondary regulatory RNA structures with FRNA viruses and has been classified into the genus *Levivivirus* (39) despite the lack of sequence similarity (27, 39) and amino acid clustering (29). The Pfam protein domain profile supports the observation that the lysis and capsid proteins of phage PP7 do not cluster to these proteins in leviviruses or the capsid and read-through proteins of alloleviviruses. However, phage PP7 replicase protein clusters with the replicase protein of the alloleviviruses but not with the leviviruses. In addition, the maturation protein of phage PP7 phylogenetically clusters with the maturation proteins of both *Levivivirus* and *Allolevivirus*.

Phage PRR1 adsorbs to host pili and displays a genetic map similar to those of viruses of the family *Leviviridae* (29). Although propagated in *Pseudomonas aeruginosa*, phage PRR1 has a broad host range (26). Phage PRR1 shared approximately 43 to 48% sequence identity to other ssRNA *Leviviridae* phages but clustered outside the *Levivivirus* and *Allolevivirus* genera (29). The PRR1 genetic map was similar to that of *Leviviridae*, and subsequently, phage PRR1 was grouped into the *Levivivirus* genus (29). Our data demonstrate that phage PRR1 shares Pfam domains with *Leviviridae* maturation, capsid, and replicase proteins. PRR1 did not share the signature *Levivivirus* 3' terminus ACCACCCA.

Phage AP205 from *Acinetobacter* shares Pfam domains in only the *Levivivirus* and *Allolevivirus* maturation and replicase proteins. However, AP205 proteins, including the coat, maturation, lysis, and replicase proteins, clustered outside the *Levivivirus* and *Allolevivirus* phylogenetic tree (29). Although AP205 lacks significant sequence similarity, this phage shares important structural features with *Leviviridae* (15). Like phages PRR1 and PP7, phage AP205 did not share the *Levivivirus* or *Allolevivirus* 3' terminus signature. NCBI GenBank taxonomy lists bacteriophage PRR1 as "unclassified *Leviviridae*," and bacteriophages PP7 and AP205 were placed into an "unclassified *Levivivirus*" category. Inclusion of these non-FRNA strains into the *Leviviridae* family should be reconsidered.

Originally, four major *Leviviridae* genogroups, I, II, III and IV, and subgroups a, b, and c in genogroup III and subgroups a and b in genogroup IV were assigned to FRNA phages based on template specificity of RNA replicase (22). To provide greater clarity in *Leviviridae* classification, we suggest designating specific phage type strains and using their names rather than alphabetical subgroups. The following phages may be

likely candidates for type strain designation: (i) phages MS2 and fr in genogroup I; (ii) phages GA and KU1 in genogroup II; (iii) phages Q β and MX1 in genogroup III; and (iv) phages NL95, SP, and FI in genogroup IV. This would result in placing the phages of genogroup I into MS2-like or fr-like categories, etc.

In conclusion, the findings of this study agree with previously determined FRNA features and phylogenetic analyses, which concluded that viruses in the family *Leviviridae* contain two genera and four distinct genogroups. Although genogroup I strain fr and genogroup III strains MX1 and M11 share only 70 to 78% sequence identity with strains in their respective genogroups, our analyses suggest that fr should be grouped into *Levivivirus* genogroup I and MX1 and M11 should be grouped into *Allolevivirus* genogroup III. Within each genus, strains share approximately 50% sequence identity, whereas between the two genera, strains have <40% nucleotide sequence identity. However, strains within each genogroup shared approximately 70 to 98% nucleotide identity.

ACKNOWLEDGMENTS

This research was funded, in part, through the Environmental Protection Agency's (EPA's) New England Regional Applied Research Effort. We gratefully acknowledge the assistance of Jack Paar III, U.S. EPA New England Regional Laboratory, for initiating and sponsoring this program.

An acknowledgment is extended to Emilie Cooper for SimPlot analysis and Syed Muaz Khalil for providing a portion of the sequence data. We thank Greg Lovelace and David Love for isolating and providing some of the strains used in this study.

The information in this document has been funded wholly (or in part) by the U.S. EPA. It has been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

This is contribution number 1354 from the Gulf Ecology Division.

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the funding agency or the Centers for Disease Control and Prevention. This article received clearance through the appropriate channels at the Centers for Disease Control and Prevention prior to submission.

REFERENCES

- Beier, H., and M. Grimm. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res.* **29**: 4767–4782.
- Bertram, G., S. Innes, O. Minella, J. P. Richardson, and I. Stansfield. 2001. Endless possibilities: translation termination and stop codon recognition. *Microbiology* **147**:255–269.
- Blattner, F. R., G. Plunkett III, C. A. Block, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome of *Escherichia coli* K-12. *Science* **277**: 1453–1462.
- Bollback, J. P., and J. P. Huelsenbeck. 2001. Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family *Leviviridae*). *J. Mol. Evol.* **52**:117–128.
- Buchen-Osmond, C. (ed.). 2006. *Levivivirus*. In ICTVdB—the Universal Virus Database, version 4. ICTVdB Management, Columbia University, New York, NY.
- Clark, A. J. 1963. Genetic analysis of a "double male" strain of *Escherichia coli* K-12. *Genetics* **48**:105–120.
- Crawford, E. M., and R. G. Gesteland. 1964. The adsorption of bacteriophage R17. *Virology* **22**:165–167.
- Engelberg-Kulka, H. 1981. UGA suppression by normal tRNA^{Trp} in *Escherichia coli*: codon context effects. *Nucleic Acids Res.* **9**:983–991.
- Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. van den Berghe, G. Volckaert, and M. Ysebaert. 1976. Complete nucleotide sequence of bacte-

- riophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**:500–507.
10. Finn, R. D., J. Tate, J. Mistry, P. C. Coghill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. L. Sonnhammer, and A. Bateman. 2008. The Pfam protein families database. *Nucleic Acids Res.* **36**:D281–D288.
 11. Friedman, S. D., E. M. Cooper, L. Casanova, M. D. Sobsey, and F. J. Genthner. 2009. A reverse transcription-PCR assay to distinguish the four genogroups of male-specific (F+) RNA coliphages. *J. Virol. Methods* **159**: 47–52.
 12. Furuse, K. 1987. Distribution of coliphages in the environment: general considerations, p. 87–124. In S. M. Goyal, C. P. Gerba, and G. Bitton (ed.), *Phage ecology*. John Wiley & Sons, Inc., New York, NY.
 13. Groeneveld, H., F. Oudot, and J. van Duin. 1996. RNA phage KU1 has an insertion of 18 nucleotides in the start codon of its lysis gene. *Virology* **218**:141–147.
 14. Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**:95–98.
 15. Klovins, J., G. P. Overbeek, S. H. E. van den Worm, H. W. Ackermann, and J. van Duin. 2002. Nucleotide sequence of a ssRNA phage from Acinetobacter: kinship to coliphages. *J. Gen. Virol.* **83**:1523–1533.
 16. Kojima, K. K., T. Matsumoto, and H. Fujiwara. 2005. Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. *Mol. Cell. Biol.* **25**:7675–7686.
 17. Loeb, T., and N. D. Zinder. 1961. A bacteriophage containing RNA. *Proc. Natl. Acad. Sci. USA* **47**:282–289.
 18. Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype c-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**:152–160.
 19. Lovmar, L., C. Fock, F. Espinoza, F. Bucardo, A. C. Syvanen, and K. Bondeson. 2003. Microarrays for genotyping human group A rotavirus by multiplex capture and type-specific primer extension. *J. Clin. Microbiol.* **41**:5153–5158.
 20. Ma, J., A. Campbell, and S. Karlin. 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**:5733–5745.
 21. Martínez, M. A., N. Verdaguer, M. G. Mateu, and E. Domingo. 1997. Evolution subverting essentiality: dispensability of the cell attachment Arg-Gly-Asp motif in multiply passaged foot-and-mouth disease virus. *Proc. Natl. Acad. Sci. USA* **94**:6798–6802.
 22. Miyake, T., I. Haruna, T. Shiba, Y. H. Itoh, K. Yamane, and I. Watanabe. 1971. Grouping of RNA phages based on the template specificity of their RNA replicases. *Proc. Natl. Acad. Sci. USA* **68**:2022–2024.
 23. Namy, O., I. Hatin, and J. P. Rousset. 2001. Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep.* **2**:787–793.
 24. Nelsen-Salz, B., H. J. Eggers, and H. Zimmermann. 1999. Integrin $\alpha_5\beta_3$ (vitronectin receptor) is a candidate receptor for the virulent echovirus 9 strain Barty. *Gen. Virol.* **80**:2311–2313.
 25. Nishihara, T., S. Fujisaki, Y. Nishimura, Y. Minami, and T. Yubisui. 2006. Analysis of six new genes encoding lysis proteins and coat proteins in *Escherichia coli* group A RNA phages. *Microbiol. Immunol.* **50**:61–66.
 26. Olsen, R. H., and P. Shipley. 1973. Host range and properties of the *Pseudomonas aeruginosa* R factor R1822. *J. Bacteriol.* **113**:772–780.
 27. Olsthoorn, R. C. L., G. Garde, T. Dayhuff, J. F. Atkins, and J. van Duin. 1995. Nucleotide sequence of a single-stranded RNA phage from *Pseudomonas aeruginosa*: kinship to coliphages and conservation of regulatory RNA structures. *Virology* **206**:611–625.
 28. Paranchych, W. 1975. Attachment, ejection and penetration stages of the RNA phage infectious process, p. 85–111. In N. D. Zinder (ed.), *RNA phages*. Cold Spring Harbor Laboratory Press, New York, NY.
 29. Ruokoranta, T. M., A. M. Grahm, J. J. Ravaniti, M. M. Poranen, and D. H. Bamford. 2006. Complete genome sequence of the broad host range single-stranded RNA phage PRR1 places it in the *Levivirus* genus with characteristics shared with alloviruses. *J. Virol.* **80**:9326–9330.
 30. Schurr, T., E. Nadir, and H. Margalit. 1993. Identification and characterization of *E. coli* ribosomal binding sites by free energy computation. *Nucleic Acids Res.* **21**:4019–4023.
 31. Skuzeski, J. M., L. M. Nichols, and R. F. Gesteland. 1990. Analysis of leaky viral translation termination codons in vivo by transient expression of improved beta-glucuronidase vectors. *Plant Mol. Biol.* **15**:65–79.
 32. Stewart, J. R., J. Vinjé, S. J. G. Oudejans, and G. I. Scott. 2006. Sequence variation among group III F-specific RNA coliphages from water samples and swine lagoons. *Appl. Environ. Microbiol.* **72**:1226–1230.
 33. Sundram, A., N. Jumanial, and M. M. Ehlers. 2006. Genotyping of F-RNA coliphages isolated from wastewater and river water samples. *Water SA* **32**:65–70.
 34. Tan, B. H., E. Nason, N. Stauber, W. Jiang, K. Monastyrskaya, and P. Roy. 2001. RGD tripeptide of bluetongue virus VP7 protein is responsible for core attachment to *Culicoides* cell. *J. Virol.* **75**:3937–3947.
 35. U.S. Environmental Protection Agency. 2004. Quality assurance/quality control guidance for laboratories performing PCR analyses on environmental samples. EPA 815-B-04-001. U.S. Environmental Protection Agency, Washington, DC.
 36. van Duin, J. 1988. The single-stranded RNA bacteriophages, p. 117–167. In C. H. Fraenkel and R. R. Wagner (ed.), *The bacteriophages*. The viruses. Plenum Press, New York, NY.
 37. van Duin, J. 1999. Single-stranded RNA phages (*Leviviridae*), p. 1663–1668. In A. Granoff and R. Webster (ed.), *Encyclopedia of virology*. Academic Press, London, United Kingdom.
 38. van Duin, J. 2000. Family *Leviviridae*, p. 645–646. In M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. S. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner (ed.), *Virus taxonomy*. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, CA.
 39. van Duin, J., and N. Tsareva. 2006. Single-stranded RNA phages, p. 175–196. In R. L. Calendar and S. T. Abedon (ed.), *The bacteriophages*. Oxford University Press, New York, NY.
 40. van Hemert, F. J., V. V. Lukashov, and B. Berkhout. 2007. Different rates of (non-) synonymous mutations in astrovirus genes; correlation with gene function. *Virol. J.* **4**:25–37.
 41. Vinjé, J., S. J. G. Oudejans, J. R. Stewart, M. D. Sobsey, and S. C. Long. 2004. Molecular detection and genotyping of male-specific coliphages by reverse transcription-PCR and reverse line blot hybridization. *Appl. Environ. Microbiol.* **70**:5996–6004.
 42. Weber, K., and W. Konigsberg. 1975. Proteins of the RNA phages, p. 51–84. In N. D. Zinder (ed.), *RNA phages*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 43. Weiner, A. M., and K. Weber. 1973. A single UGA codon functions as a natural termination signal in the coliphage Q- β coat protein cistron. *J. Mol. Biol.* **80**:837–855.
 44. Zinder, N. D. 1965. RNA phages. *Annu. Rev. Microbiol.* **19**:455–473.